systems

MDPI

Article

An IoT-Platform-Based Deep Learning System for Human Behavior Recognition in Smart City Monitoring Using the Berkeley MHAD Datasets

Othman O. Khalifa [1,2], Adil Roubleh [1], Abdelrahim Esgiar [3], Maha Abdelhaq [4], Raed Alsaqour [5,*], Aisha Abdalla [1], Elmustafa Sayed Ali [6,7] and Rashid Saeed [8]

[1] Department of Electrical and Computer Engineering, Kulliyyah of Engineering, International Islamic University Malaysia, Kuala Lumpur 50728, Malaysia
[2] Libyan Centre for Engineering Research and Information Technology, Bani Waleed 411, Libya
[3] Department of Electrical and Electronic Engineering, Faculty of Engineering, Sirte University, Sirte 674, Libya
[4] Department of Information Technology, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia
[5] Department of Information Technology, College of Computing and Informatics, Saudi Electronic University, P.O. Box 13316, Riyadh 93499, Saudi Arabia
[6] Department of Electrical & Electronics Engineering, Faculty of Engineering, Red Sea University, Port Sudan 36481, Sudan
[7] Department of Electronics Engineering, Faculty of Engineering, Sudan University of Science and Technology (SUST), Khartoum 00407, Sudan
[8] Department of Computer Engineering, College of Computers and Information Technology, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia
* Correspondence: r.alsaqor@seu.edu.sa

Citation: Khalifa, O.O.; Roubleh, A.; Esgiar, A.; Abdelhaq, M.; Alsaqour, R.; Abdalla, A.; Ali, E.S.; Saeed, R. An IoT-Platform-Based Deep Learning System for Human Behavior Recognition in Smart City Monitoring Using the Berkeley MHAD Datasets. *Systems* **2022**, *10*, 177. https://doi.org/10.3390/systems10050177

Academic Editor: Paolo Visconti

Received: 27 August 2022
Accepted: 27 September 2022
Published: 1 October 2022

Abstract: Internet of Things (IoT) technology has been rapidly developing and has been well utilized in the field of smart city monitoring. The IoT offers new opportunities for cities to use data remotely for the monitoring, smart management, and control of device mechanisms that enable the processing of large volumes of data in real time. The IoT supports the connection of instruments with intelligible features in smart cities. However, there are some challenges due to the ongoing development of these applications. Therefore, there is an urgent need for more research from academia and industry to obtain citizen satisfaction, and efficient architecture, protocols, security, and services are required to fulfill these needs. In this paper, the key aspects of an IoT infrastructure for smart cities were analyzed. We focused on citizen behavior recognition using convolution neural networks (CNNs). A new model was built on understanding human behavior by using the berkeley multimodal human action (MHAD) Datasets. A video surveillance system using CNNs was implemented. The proposed model's simulation results achieved 98% accuracy for the citizen behavior recognition system.

Keywords: smart cities; Internet of Things; neural networks; video surveillance; deep learning; artificial intelligence; Berkeley MHAD Datasets

1. Introduction

Smart city monitoring aims to improve people's quality of life and the performance of services by using the latest technology [1]. Data are acquired from the many devices that serve in the smart cities and include data such as videos, security surveillance, environment, e-government, transportation, etc. The IoT involves intelligent devices, sensors, wireless devices such wireless sensors, and radio-frequency identifications (RFIDs) built into service systems, being connected in a network [2]. Data are collected from devices classified and used for decision making. Many smart city applications have been introduced such as transportation [3], healthcare, environment monitoring, public safety [4], and many others. Therefore, leveraging CNN and machine learning (ML) can be used

in smart city monitoring [5–9]. These techniques aim to develop algorithms that can process input data for learning and accordingly be able to predict unknown information or actions. These algorithms could be categorized into two streams, supervised learning and unsupervised learning.

Supervised learning enables one to find a certain mapping to predict the outputs of unknown data [10]. Unsupervised learning focuses on exploring the intrinsic characteristics of inputs. Since supervised learning leverages the labels of inputs that are understandable to a human, it can apply to pattern classification and data regression problems [11,12]. However, supervised learning relies on labeled data, which needs a considerable amount of manual work. Moreover, there could be uncertainties and ambiguities in labels as well [13–16]. Additionally, the label for an object is not unique. To tackle these problems, unsupervised learning can be used to handle the intra-class variation, as it does not require the labels of data. In previous research, ML techniques were applied in many applications such as computer vision, bioinformatics, medical applications, natural language processing, speech processing, robotics, and stock market analysis [17]. The use of the deep learning (DL) approach enables the improvement of the detection and recognition processes in IoT-based human recognition platforms [18,19]. Accordingly, we proposed a new model for citizen behavior recognition based on IoT infrastructure for smart city monitoring. The model uses convolution neural networks (CNNs) with the Berkeley MHAD Datasets, which helps to understand human behavior and improves the performance of designed video surveillance systems with high accuracy.

The outcome of this study is to improve human behavior modeling for various smart city applications. The existing smart city systems use human behavior to promote smart agents. The proposed model helps to build a knowledge base of human behavior from various sources using sensory data and IoT technologies.

The main motivation of this research was to detect suspicious human behavior, which will help to identify activities such as fighting, slapping, vandalism, and people running in public places, schools, or colleges [5]. This paper focuses on the recognition of some actions including jumping, jumping jacks, boxing, waving two hands, waving one hand (the right hand), and clapping hands. The novelty of this paper can be summarized in the following two points:

- We developed a new framework for modeling human behavior-based deep learning models to understand and analyze human behavior better. The proposed algorithms explore convolution deep neural networks, which learn different features of historical data to determine collective abnormal human behaviors;
- We tested and evaluated the experiments of human behavior recognition systems based on convolution deep neural networks to demonstrate the usefulness of the proposed method.

An IoT-platform-based deep learning system for human behavior recognition is of benefit to society and the industry of smart city monitoring. Some high-value services and applications may involve:

- Safety and security services, i.e., suicide deterrence in municipal places, amenability monitoring, and the scrutiny of disaster mitigation due to the detection of vandalism in a crowd, the protection of critical infrastructures, the detection of violent and dangerous situations, perimeter monitoring and person detection, and weapon detection and reporting;
- Epidemic control policy services, i.e., social distancing in municipal spaces, automatic mask recognition, sanitary compliance detection, and monitoring healthcare;
- Infrastructure and Traffic monitoring, i.e., monitoring traffic in smart cities, the recognition of traffic rule violations, the surveillance of roadsides, and parking space management.

The paper is organized as follows: Section 2 provides an overview of the background and related work. Section 3 presents the proposed video surveillance system. The experi-

mental results and discussion are illustrated in Section 4. Result validation is provided in Section 5, and finally, Section 6 provides the conclusions and future work.

2. Background and Related Work

2.1. Smart Sustainable Cities

Smart sustainable cities are innovative cities that use information and communication technologies (ICTs) to improve the quality of life, the efficiency of operations and services, and competitiveness while ensuring that the needs of the present and future generations concerning the economic, social, environmental, and cultural aspects are met. Smart cities have emerged as a possible solution to the problems related to sustainability that result from rapid urbanization [20]. They are considered imperative for a sustainable future. In general, those cities that aim to become smart sustainable cities have to become more attractive, sustainable, inclusive, and more balanced for the citizens who live or work in them, as well as city visitors. Figure 1 shows the classification of some applications in smart cities [21].

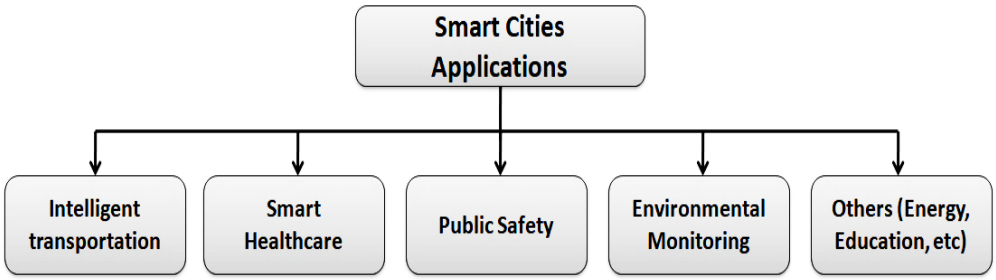

Figure 1. Smart city applications.

Smart cities have become a focus for many governments globally, as citizens demand satisfaction in terms of efficient services and smooth, secure transportation [22]. The concept of the sustainability of smart cities falls within the scope of data and information sustainability. Sustainable urban development needs clean and transparent data that represent the different aspects of smart cities and are available at an individual level. Furthermore, these data must be freely used for data exchange in IoT networks [23].

The concept of smart sustainable cities is linked to the possibility of obtaining the right information at the right time to help in making decisions by citizens or government service providers to improve the quality of life [24]. The process of monitoring abnormal human behavior through IoT platforms helps to improve the intelligent management of smart cities, which enables the transformation of the social behavior of citizens toward the sustainability of city resources by making decisions to produce new smart city management standards and rules [25]. It also helps the government evaluate citizens' behavior to improve services, in addition to environmental, social, and economic sustainability.

Artificial intelligence (AI) techniques help to improve the performance of smart sustainable cities that are based on IoT networks. These technologies offer effective solutions in intelligent transportation, urban planning [26], data confidentiality, and big data processing. It also helps in decision-making processes and predicting possible future events. AI technologies are involved in many smart city applications providing solutions for traffic congestion, energy data analysis, health care diagnostics, and cyber security [27]. Some examples are shown in Figure 1.

2.2. IoT-Platform-Based Deep Learning Systems

Deep learning (DL) is a part of artificial neural networks (ANNs). It is a computation model inspired by biological principles from the human brain. This field has been studied extensively for decades. The ANN is composed of connected artificial neurons that simulate the neurons in a biological brain. The weights between the layers of neurons are based on a non-linear transformation function called sigmoid [28]. The main objective of ML is

to develop algorithms that are capable of learning and making accurate predictions in a given task.

In recent years, DL achieved a noteworthy achievement in computer vision. The creators of AlexNet accomplished a record in the execution of a profoundly difficult dataset named ImageNet [29]. AlexNet was capable of ordering millions of high-resolution images from different classes with the best blunder rate. DL approaches are ML methods that work on numerous (multi-layer) dimensions [30]. Convolutional neural networks (CNNs or ConvNets) are considered to be the most important DL architecture for human action recognition [31].

Different DL structures have been previously proposed and appear to produce state-of-the-art results on numerous assignments, not limited to human activity recognition [32]. As one of the most crucial deep learning models, CNNs obtain superior results in solving visual-related tasks. A CNN is a type of artificial neural network, intended for processing visual and other two-dimensional information. The main advantage of this model is that it works specifically on crude information with no manual feature extraction. The CNN model was first introduced in 1980 by Fukushima [33]. CNNs are inspired by the structure of the visual nervous system [34]. CNN models continue to be proposed and developed.

Figure 2 shows a summarized statistic about the number of articles focused on smart city applications based on AI, ML, and DL technologies from 2018 to September 2022. Deep learning (DL) techniques help to clarify innovative solutions that deal with the challenges facing smart applications in urban cities related to the environment, people, transportation, and security. These technologies help to improve data processing, transform data into useful information, and help develop cognitive intelligence for sustainable cities [35]. Recently, the combination of ML and DL techniques has become more common as an unsupervised neural network and has been used to identify patterns and objects through videos in many applications, providing high detection ability with an accuracy level of more than 80%. DL technologies analyze the aggregated and integrated big data including images, videos, sensors, cloud computing, and resource management mechanisms to implement many intelligent operations related to detection and prediction [36].

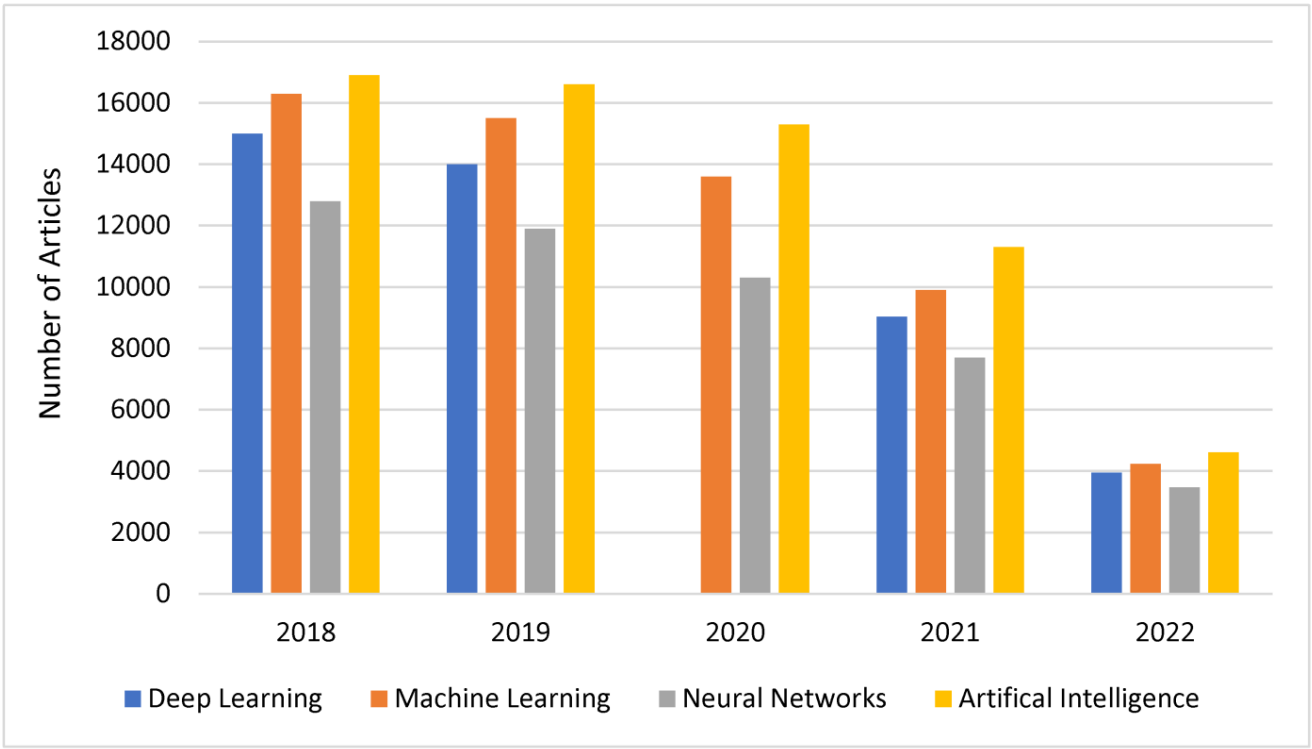

Figure 2. Deep learning (DL) approach in smart cities—research analysis (Google Scholar).

2.3. Related Work

Several studies have recently been presented to monitor human behavior using deep learning mechanisms for different applications [37]. Our proposed system is focused on designing an IoT-platform-based DL system for citizen behavior recognition. The proposed algorithm is a part of an IoT-based citizen behavior recognition project that integrates the structure of smart cities based on the IoT with the system of monitoring human behavior and predicting suspicious events, which helps the competent authorities to take appropriate actions. From previous studies, we found only a few models that address such a study, which depend on the mechanisms of deep learning to investigate human behavior. By reviewing these studies, the differences between them and our study are presented in Table 1. The previous studies considered different methodologies for human behavior recognition with high detection performance; however, most of them give lower accuracy than our proposed method.

Table 1. Comparison between previous studies and the proposed method.

Citations	Methodology	Datasets	Max Accuracy	Shortcomings Compared with Proposed Study
Shanshan et al., 2019 [20]	Human-behavior-recognition-based DL for the posture of the human body	UCI HAR dataset	93%	Large testing error and low accuracy
Rashmi et al., 2020 [21]	Human-action-recognition-based CNN to extract features from skeleton joint information	MSRAction3D dataset	97%	Not recognizing different human motions. It uses distance features which need more computations
Nirmalya et al., 2021 [22]	Human behavior detection during activities of daily living using the EDSCCA algorithm	ADLs dataset	83.87%	Not recognizing different motions and low accuracy
Xiwei Liu, 2022 [23]	Development of 3D residual structures for recognizing human behavior using the DL approach	HMDB51 and UCF101 datasets	80%	The difficulty of training and low accuracy
Palash et al., 2022 [24]	Detection carried objects by humans based on CNN	ImageNet, Open Image dataset, and Olmos dataset	97.5%	Detects only objects carried by humans, and now motion detection

3. Human Behavior Recognition Methodology

After investigating all the recently developed deep learning architecture, a recurrent neural network was selected based on the many needed and required features for recognizing human activities as well as the pattern recognition used throughout this work. The proposed solution consisted of two main phases, as shown in Figure 3. These phases are testing and training, which describe the steps of video processing based on the HAR approach. In the testing phase, the video input is captured with a high-resolution video sensor, taking a set of image samples, and then passed to the image preprocessing unit, which is used to convert the analog-captured video to normalized datasets and enhance the quality of feature extraction for analysis preparation.

The extracted datasets are then used as input into the human detection and segmentation unit, which enables the recognition of the sample of video objects based on the intelligent model approach used. The dimensions of the segmented samples are then reduced by efficiently representing a large number of pixels from the sample using the feature extraction and representation unit to effectively capture parts of interest [31]. Then, the extracted features of the dataset are trained using the DL and RNNS algorithms and matched with the filtered detected pattern of the captured video in the training phase to output the human behavior recognition result.

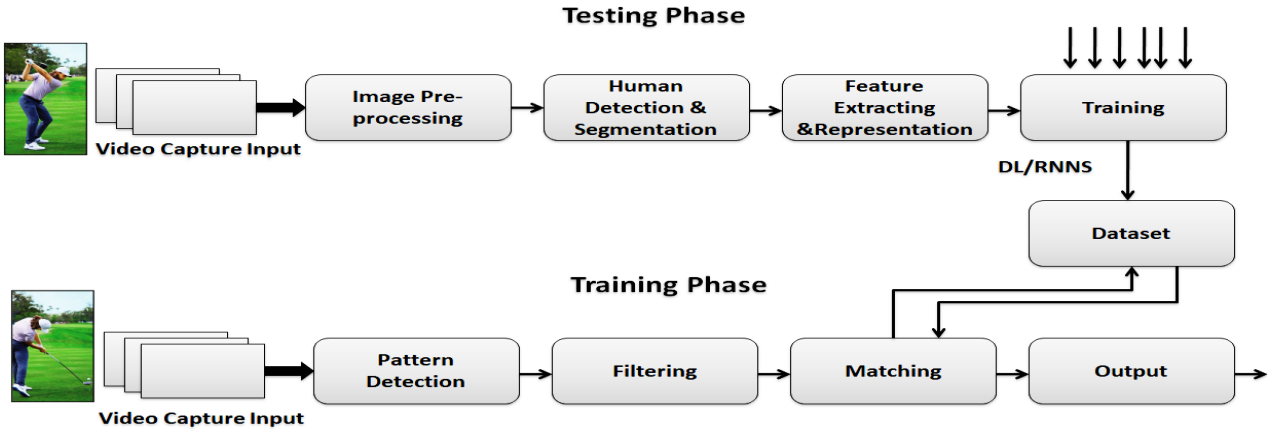

Figure 3. The proposed solution for human behavior recognition.

Different DL recognition structures have been previously proposed and appear to produce state-of-the-art results on numerous tasks. Overall, DL approaches are ML methods that work on numerous (multi-layer) dimensions [34]. Furthermore, the data were preprocessed to yield a skeleton of the body part to be detected in an activity, to reduce the number of incorrect predictions produced by the model. One of the main problems for human activity recognition (HAR) is the view-invariant issue, wherein the model is only able to detect activities from the same viewpoint that it was trained to detect. However, integrating deep learning with pose estimation technology should overcome this challenge.

The proposed DL recognition structure is shown in Figure 4. The hierarchical model enables the extraction of video features using a classification task. It consists of many layers that are used to represent the location and direction, makes a combination with corresponding objects, detects familiar objects in the video, and obtains recognition results [26]. The deep learning network obtains the hierarchical recognition tasks. The process of human behavior recognition (HBR) is described in Algorithm 1.

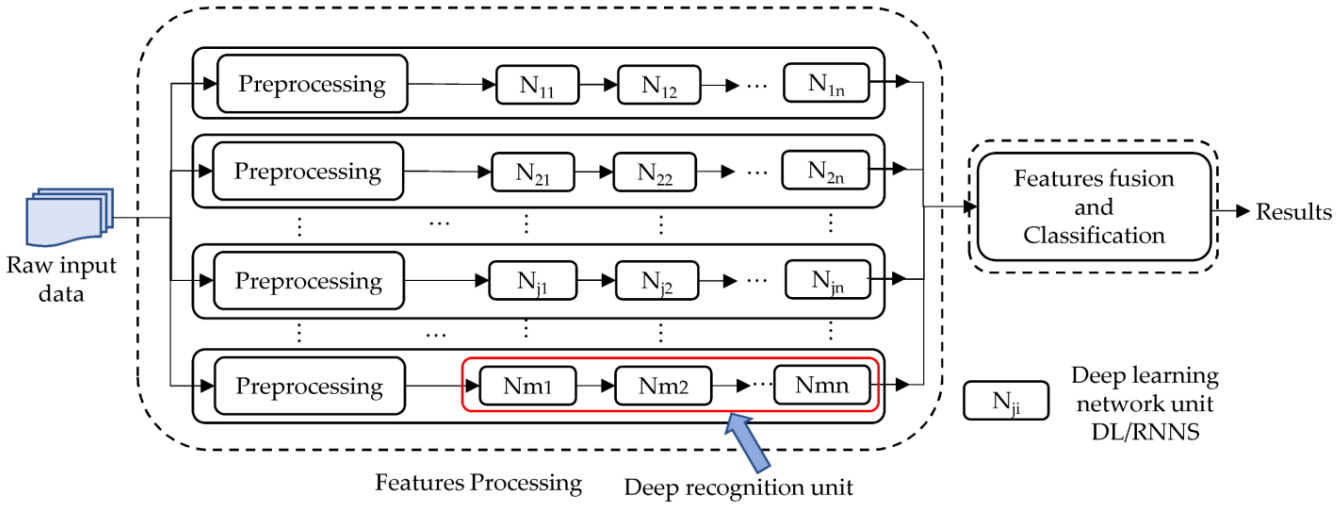

Figure 4. Proposed deep recognition Architecture.

In Algorithm 1, we developed a new method for HBR based on deep learning to study and analyze certain HB aspects. The proposed algorithm explores CDN, which learns different historical data features. To test and evaluate Algorithm 1, we applied it to different experiments for HBR systems based on CDN networks to demonstrate the effectiveness of the proposed procedure.

Firstly, we began with training and testing, during which the signal acquisition was set and preprocessing was conducted. Secondly, the algorithm calculated the extracted features; the global and local features were identified. Thirdly, by using the extracted global and local features, the baseline was set. Fourthly, we tested the process and used the global and local features to compare with the baseline and state the classification model. Finally, the recognition mode was activated and performed.

The HBR enables the processing of the captured video by dividing it into several samples, where each sample is separately processed in a multi-layer structure. All deep learning modules extract information for segments in parallel, so that the features can be extracted without more redundant information [26]. In this model, each deep learning unit is jointly trained, which adds greater advantages to the process of identifying events and human behaviors. To extract more features from the captured video samples, the deep learning units were trained on a set of different exercises at the same time with the samples from three categories related to dynamic and static samples in addition to selectivity for feature extraction [27]. These three categories enabled us to detect human jumping, boxing, waving, and clapping.

Algorithm 1. Human Behavior Recognition (HBR) Algorithm

Initiate training process
Initiate Testing process

1. for each training and testing
2. **Set**: signal acquisition
3. **Do**: preprocessing
4. **Calculate:** and extract features
5. **Set:** global and local features
6. for the training process
7. if (***global and locate***) features extracted
8. **Set:** baseline
9. end
10. end
11. for the testing process
12. if (***global and locate***) features extracted
13. **Do:** comparison with output in *step 8*
14. **Make:** a classification model
15. ***Recognize action***
16. end
17. end

The process was performed in four main stages. First, the input video was obtained from the Berkeley MHAD Datasets [16,17] which have been preprocessed using an open-pose bottom-up approach. Second, the data were processed through human detection and segmentation, using filters and patterns to find the overall commonalities in those patterns, as previously described. Feature extraction and representation were used to select an action from the dataset by going deeper into the compression to where the patterns were assembled [38]. Action recognition was the last step, where the action was detected and extracted from a known pattern class.

4. Experimental Results and Discussion

The proposed CNN model consisting of two convolution layers, two maximum pooling layers, and two full connection layers is shown in Table 2. The active function used rectified linear unit (ReLU), and the input image was converted into a 28×28 size monochrome image. The model was built using the recurrent neural network (RNN) architecture. The parameters that could vary and thus change the accuracy and the precision of the overall activity recognition were the numbers of epochs, the batch size, and the iterations. A single epoch was a point at which an entire dataset was passed forward and in reverse through the neural system just once. In contrast, the batch size was the number of training samples

in one batch. Iterations were defined as the number of batches needed to complete one epoch. As mentioned earlier, the dataset was the Berkeley MHAD Datasets [39], which contains 11 activities. Six of them were used in this model, including jumping, jumping jacks, boxing, waving two hands, waving one hand (the right hand), and clapping hands.

Table 2. Details of the CNN model.

Name	Size	Function
training_data_count	test_data_count	4519 training series (with 50% overlap between each series)
test_data_count	len(X_test)	1197 test series
n_input	len(X_train [0][0])	Number of input parameters per timestep
Hidden layer	34	Hidden layer number of features
No. of classes	6	Number of classes
Decaying learning rate	True	Calculated as: decayedlearningrate = learningrate * decayrate ^ (globalstep / decaysteps)
Learning rate	0.0025	Used if decaying learning ate set to false
Initial learning rate	0.005	A starting point for learning rate.
Decay rate	0.96	The base of the exponential in the decay
Decay steps	128,256,512	Every 60,000 steps with a base of 0.96
Global step	tf.Variable(0, trainable = False)	The parameter in the learning rate pushes it to take another step in the learning process.
Training iterations	training_data_count 100,200,600,1000	Loop 100,200,600,1000 times on the dataset, i.e., 100,200,600,1000 epochs
Batch size	128	Number of training samples present in a one batch
display_iter	batch_size X 8	To show test set accuracy during training

The best result, illustrated in Figure 5, showed an accuracy of ~99% and a precision of ~99%. Furthermore, the figure compares the accuracy of testing and training. In training, the accuracy obtained 100% in some cases, so the overall test accuracy reaching 99% was a good result. The rows in the normalized confusion matrix are the actual class of activities, and the columns are the predicted class of activities. Thus, the diagonal elements represent the degree of correctly predicted classes.

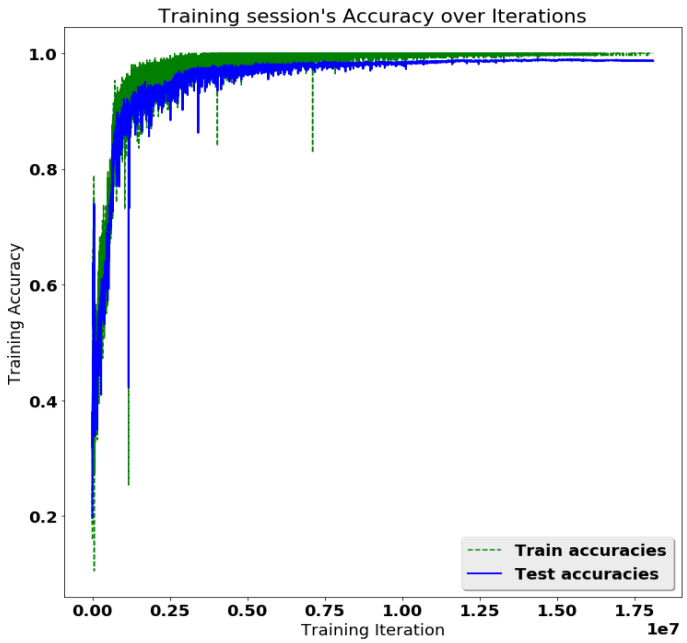

Figure 5. Comparison of the training session's accuracy of testing over training iteration.

To visualize the algorithm's performance in the six classes, a confusion matrix was used, as shown in Figure 6. For this model, the batch size value was 256, and the number of epochs was 800. In addition, as shown in Figure 6, it was observed that, due to color darkening, there was a slight similarity between clapping hands and boxing, and boxing with waving one hand, which is understandable, as these activities have much in common. The testing accuracy was 98.69%.

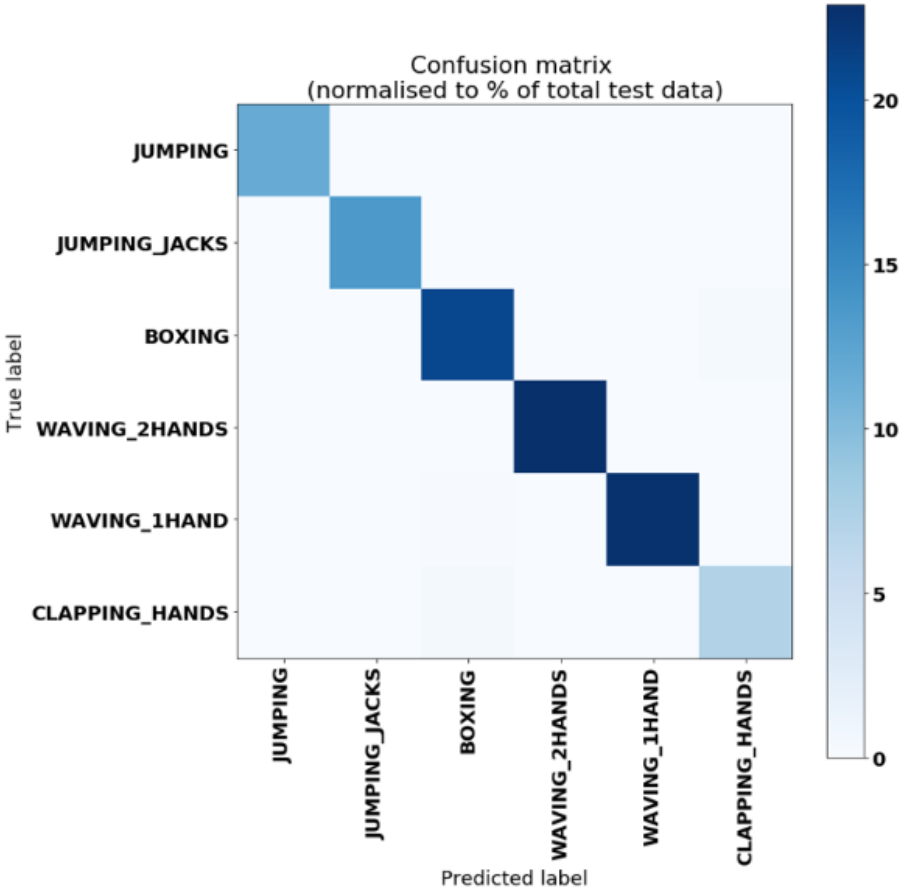

Figure 6. The confusion matrix of the proposed algorithm's performance on the six classes.

Figure 7 shows that the overall accuracy dropped to 97%, and Figure 8 shows its confusion matrix. Theoretically, fewer epochs should cause a drop in the overall accuracy, as the level of learning would still be high, and there would still be many factors on which the model needs to be trained. This is commonly known as underfitting. This needs to be verified experimentally. The batch size was set to 256, the same as the previous one, but the number of iterations was reduced from 800 to 600 epochs.

Figure 9 shows the result and confirms that the overall accuracy drastically dropped to 93%. Theoretically, a larger batch size would cause a drop in the overall accuracy. The verification of this was required. The batch size was increased to 512, and the iterations were reduced to 100 epochs to reduce the training time. Furthermore, as can be seen from the confusion matrix in Figure 10, there was confusion between boxing and waving with one hand, and boxing and clapping, in addition to waving with two hands and jumping jacks.

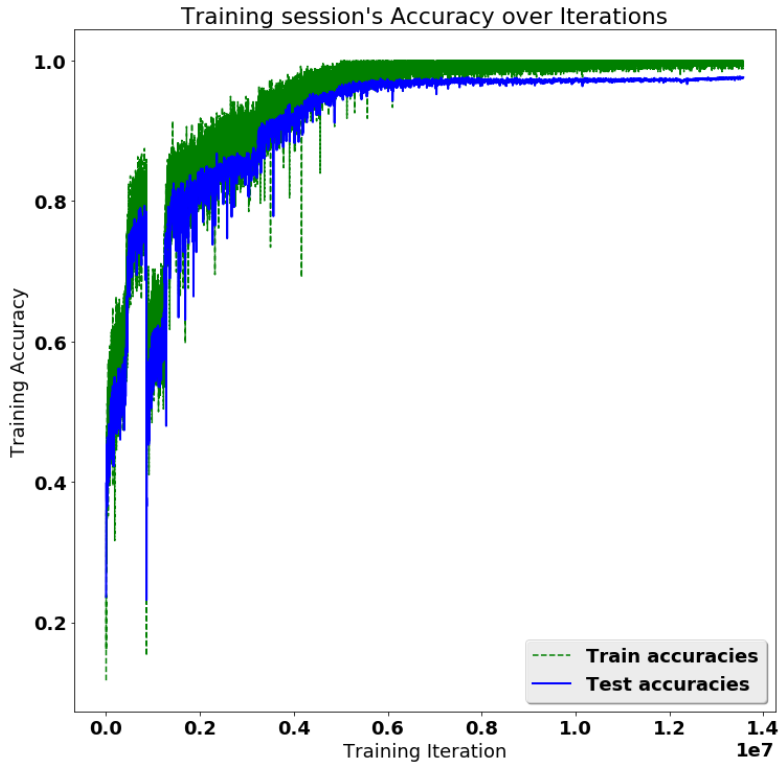

Figure 7. Training sessions accuracy for testing a model with a 256-batch size and 600 epochs.

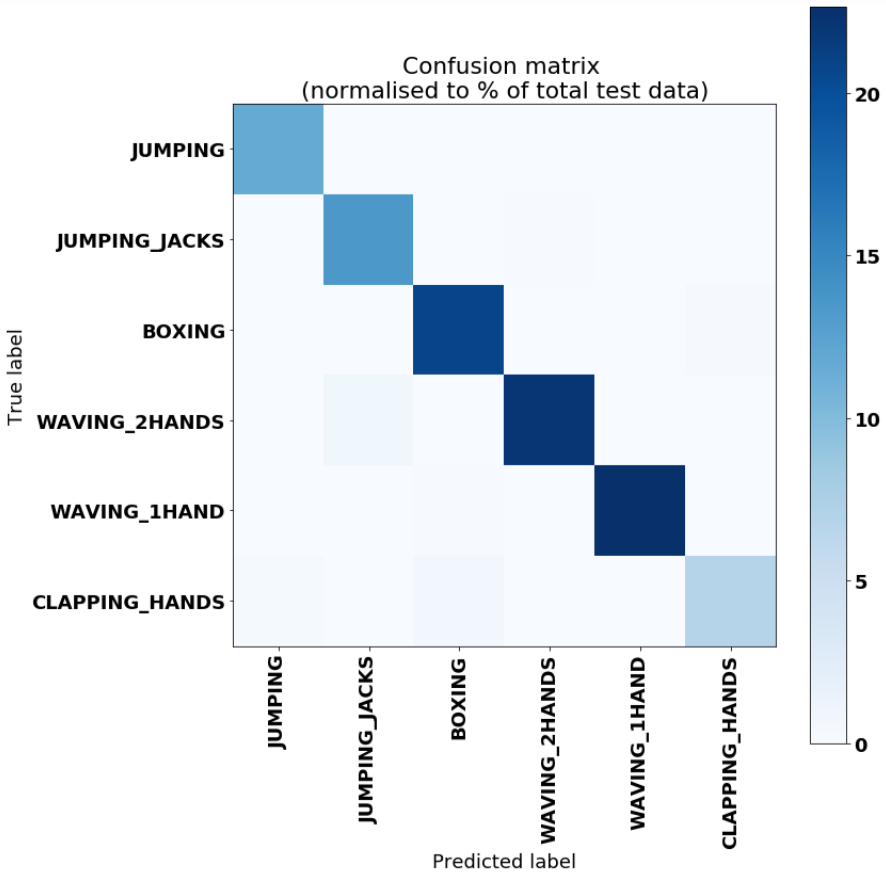

Figure 8. The confusion matrix for the model with a 256-batch size and 600 epochs.

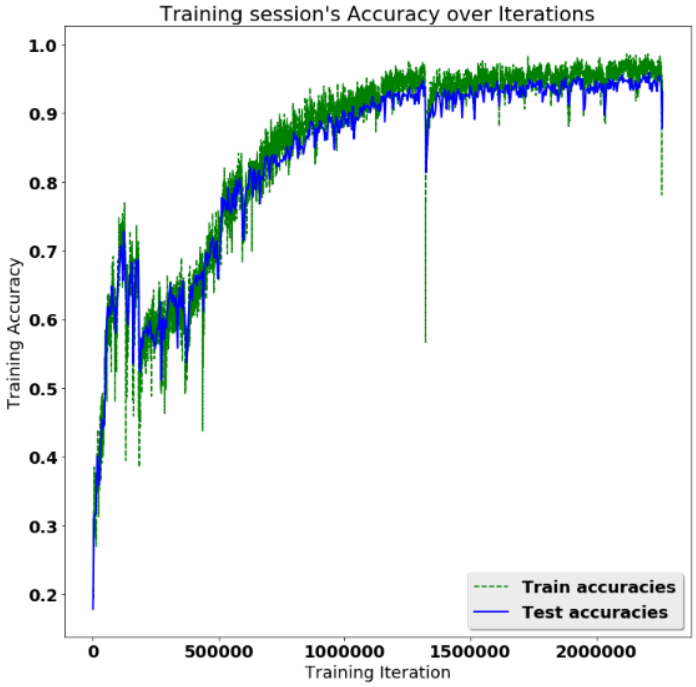

Figure 9. Training session's accuracy results from testing a model with a 512-batch size and 100 epochs.

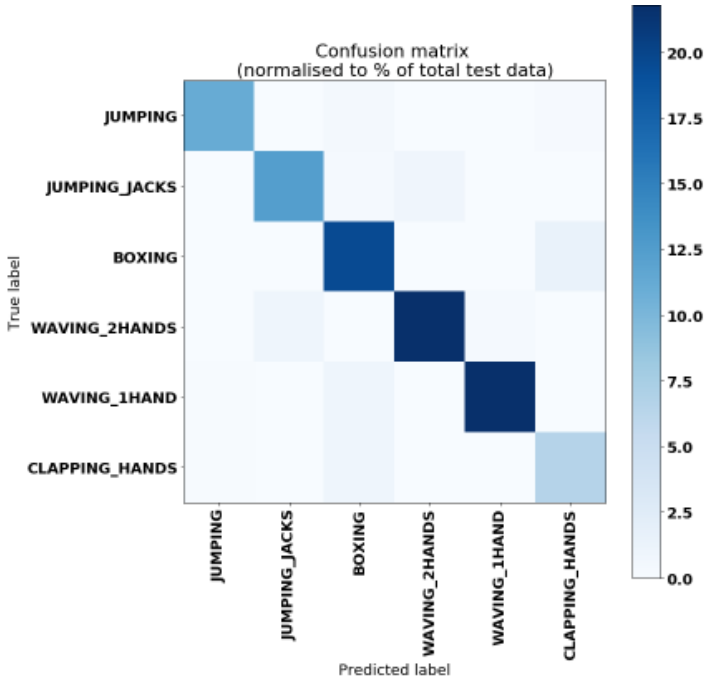

Figure 10. The confusion matrix for a model with a 512-batch size and 100 epochs.

Another experiment was conducted where the batch size was less than 256, as shown in Figure 11. The batch size was 128, and the number of epochs was 1000. The result for the overall accuracy was 97.91%, as shown in Figure 12. Unnecessarily increasing the number may cause propagation in the signals, creating an error in some values. The batch size of 128 with fewer iterations provided the best result thus far.

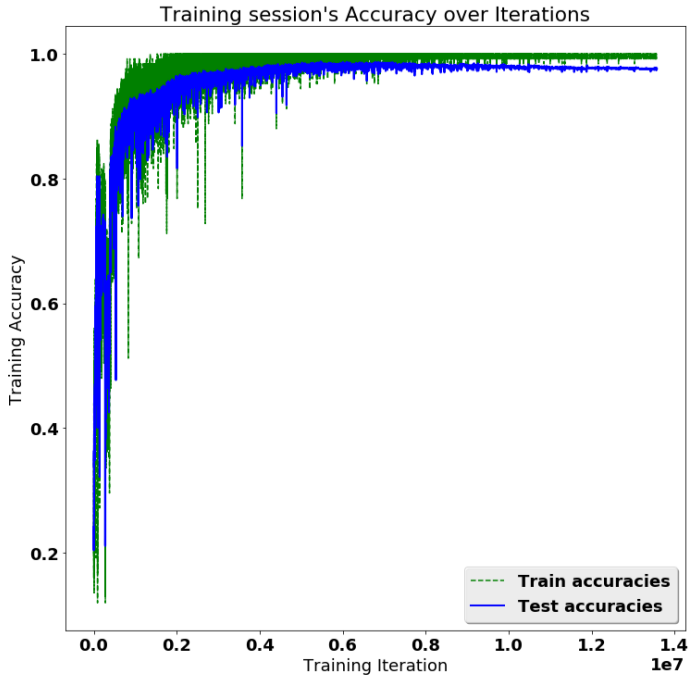

Figure 11. Result of testing a model with a 512-batch size and 1000 epochs.

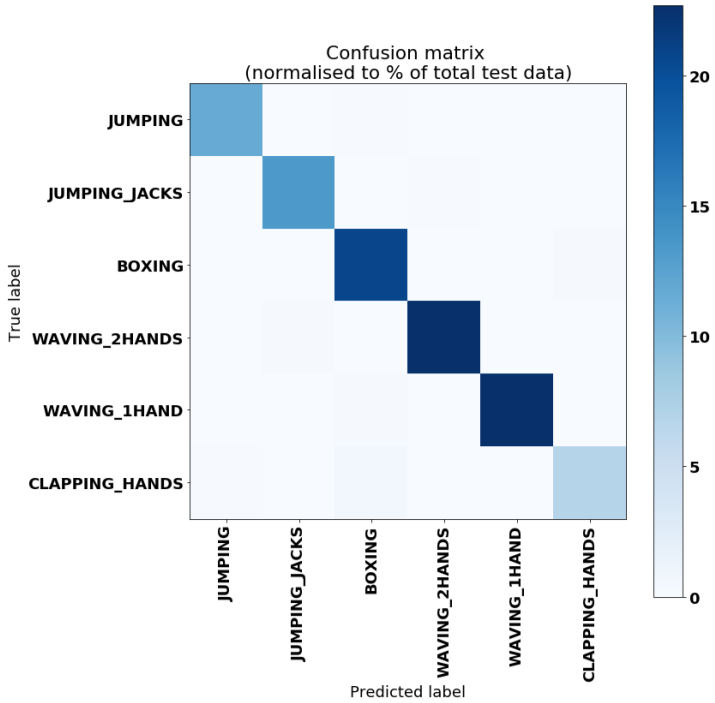

Figure 12. The confusion matrix for a model with a 128-batch size and 1000 epochs.

Figure 13 shows that the model batch size was decreased to 128, as discussed earlier, with fewer iterations, such as 200 epochs, which was enough to train the model without causing signal propagation. Figure 13 shows the result was 98.71%, which was as expected and was the highest result obtained thus far. The trend shown in the test exponentially closed on the training trend more than in previous models. Moreover, the confusion matrix was the clearest obtained thus far, as shown in Figure 14.

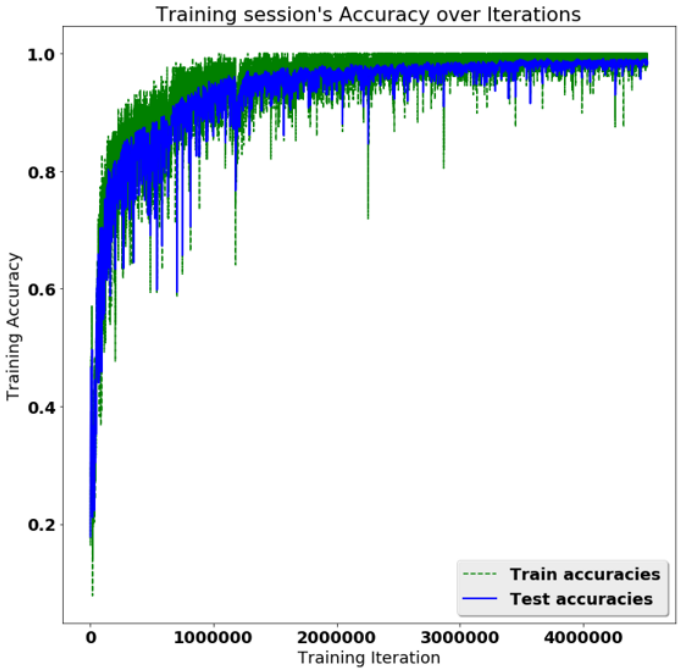

Figure 13. Training session's accuracy results from testing a model with a 128-batch size and 200 epochs.

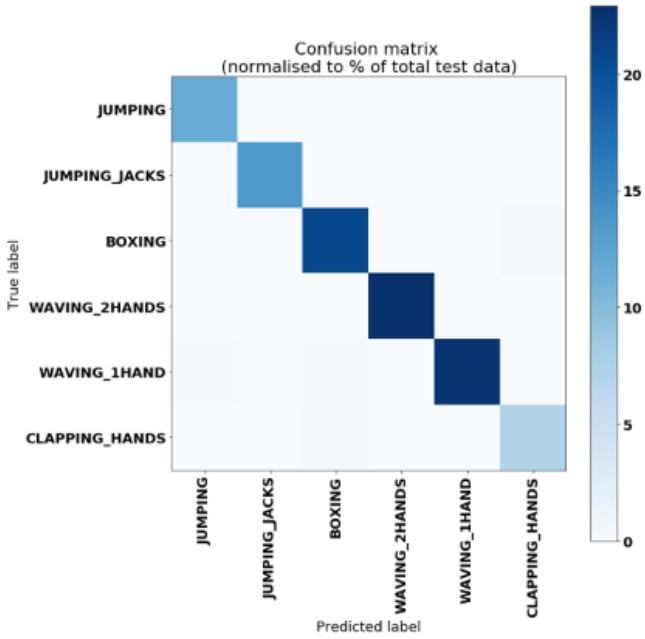

Figure 14. The confusion matrix of a model with a 128-batch size and 200 epochs.

The above results, shown in Figures 6–14, indicate that when the model batch size was decreased, fewer iterations were enough to train the model without causing signal propagation. This practically confirmed that a larger batch size caused a drop in the overall accuracy. It was also shown that unnecessarily increasing the number of epochs may cause propagation in the signals, creating an error in some values.

Previously, many tests have been conducted between the values of X_train, Y_train, and the X_test and Y_test. The dataset was split with an allocation of 80% for training and 20% for testing. Figures 7, 9, 11 and 13 present the test results. The view-invariant issue required using a preprocessed dataset that used the open-pose method to obtain body parts

as the input to our model. This trained the model to recognize the action regardless of the view with an accuracy level of ~99%. Table 3 shows the results when the parameters were changed.

Table 3. A summary of results using different parameters.

Batch Size	Number of Epochs	Overall Test Accuracy
512	300	93.45%
512	300	89.688%
128	200	98.72%
128	200	98.48%
128	300	98.38%
128	1000	97.90%
128	150	97.00%
256	800	98.69%
512	100	93.16%
512	600	97.61%
512	100	97.92
512	150	96.43%

5. Result Validation

To validate that the model was working with high accuracy and that it could detect actions from different viewpoints, a video was uploaded that was not from the training dataset. This was carried out to check if the model could identify the activity's label. Figure 15 shows the images from the uploaded video.

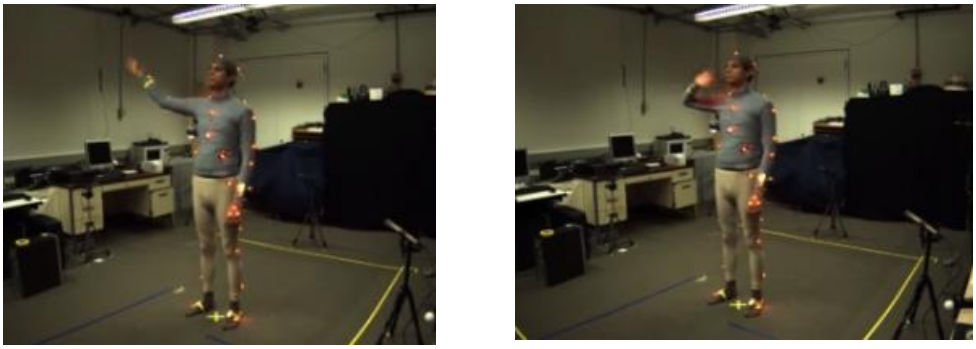

Figure 15. These images show the activity of waving with one hand (right hand).

The video was converted to an array suitable to be uploaded in the Jupyter notebook. The result of the predicted activity is shown in Figure 16.

```
#IMG_SIZE= 70
#Img_array=cv2.imread(filepath,cv2.IMREAD_GRAYSCALE)
#new_array=cv2.resize(img_array,(IMG_SIZE,IMG_SIZE))
predicted_activity= LABELS[int (one_hot_predictions[0][0])]
print ("The name of this activity is :",predicted_activity)
```

```
The name of this activity is : WAVING_1HAND
```

Figure 16. The resulting array of the activity prediction.

6. Conclusions

Smart sustainable city monitoring is challenging for many reasons, such as different application domains requiring different tasks. Furthermore, a large volume of data of

different types and modalities requires different algorithms and analysis techniques. This paper proposed and simulated an IoT platform for human behavior recognition using CCNs. The development of citizen behavior and activity recognition may start by overcoming a small issue, such as a view-invariant problem. As currently used, they perform poorly when tested with real-life scenario viewpoints and complex tasks.

The study showed that there are still many challenges ahead for this emerging technology owing to the complex nature of the deep and wide coverage of smart city applications. A model was built on understanding simple human behaviors such as jumping, jumping jacks, boxing, waving two hands, waving one hand (the right hand), and clapping hands. A video surveillance system using CNNs was implemented. The simulation results showed the overall accuracy was about 98%. Future work could investigate suspicious human activity such as fighting, slapping, vandalism, and people running in public places, both indoors and outdoors.

Author Contributions: Conceptualization, A.R. and A.E.; methodology, O.O.K. and A.R.; software, A.R.; validation, O.O.K. and A.A.; formal analysis, A.R. and A.E.; investigation, E.S.A.; resources, R.A.; data curation, E.S.A. and R.S.; writing—original draft preparation, A.R. and E.S.A.; writing—review and editing, M.A. and R.A.; visualization, A.R. and A.E.; supervision, O.O.K. and A.A.; project administration, R.S.; funding acquisition, M.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project Number (PNURSP2022R97), and Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: Princess Nourah bint Abdulrahman University Researchers Supporting Project Number (PNURSP2022R97), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Saeed, R.A.; Mabrouk, A.A.M.H.; Mukherjee, A.; Falcone, F.; Wong, K.D. WiMAX, LTE, and WiFi Interworking. *J. Comput. Syst. Netw. Commun.* **2010**, *2010*, 754187. [CrossRef]
2. Bokhari, S.A.A.; Myeong, S. Use of Artificial Intelligence in Smart Cities for Smart Decision-Making: A Social Innovation Perspective. *Sustainability* **2022**, *14*, 620. [CrossRef]
3. Ghazal, T. A review on security threats, vulnerabilities, and countermeasures of 5G enabled Internet-of-Medical-Things. *IET Commun.* **2022**, *16*, 421–432.
4. Pradeep, P.; Kant, K. Conflict Detection and Resolution in IoT Systems: A Survey. *IoT* **2022**, *3*, 12. [CrossRef]
5. Ali, E.S.; Hassan, M.B.; Saeed, R.A. Machine Learning Technologies in the Internet of Vehicles. In *Intelligent Technologies for Internet of Vehicles*; Magaia, N., Mastorakis, G., Mavromoustakis, C., Pallis, E., Markakis, E.K., Eds.; Springer: Cham, Switzerland, 2021; pp. 225–252.
6. Poulter, A.J.; Cox, S.J. Enabling Secure Guest Access for Command-and-Control of Internet of Things Devices. *IoT* **2021**, *2*, 13. [CrossRef]
7. Hassan, M.B.; Ali, E.S.; Mokhtar, R.A.; Chaudhari, B.S. 6—NB-IoT: Concepts, applications, and deployment challenges. In *LPWAN Technologies for IoT and M2M Applications*; Chaudhari, B.S., Zennaro, M., Eds.; Academic Press: Cambridge, MA, USA, 2020; pp. 119–144.
8. Ali, E.S.; Hasan, M.K.; Hassan, R.; Hassan, M.B.; Islam, S.; Nafi, N.S.; Bevinakoppa, S. Machine Learning Technologies for Secure Vehicular Communication in Internet of Vehicles: Recent Advances and Applications. *Secur. Commun. Netw.* **2021**, *2021*, 8868355. [CrossRef]
9. Ghorpade, S.N.; Zennaro, M.; Chaudhari, B.S.; Saeed, R.A.; Alhumyani, H.; Abdel-Khalek, S. A novel enhanced quantum PSO for optimal network configuration in heterogeneous industrial IoT. *IEEE Access* **2021**, *9*, 134022–134036. [CrossRef]
10. Pappalardo, M.; Virdis, A.; Mingozzi, E. An Edge-Based LWM2M Proxy for Device Management to Efficiently Support QoS-Aware IoT Services. *IoT* **2022**, *3*, 11. [CrossRef]
11. Vashisht, M.; Kumar, B. Effective Implementation of Machine Learning Algorithms Using 3D Colour Texture Feature for Traffic Sign Detection for Smart Cities. *Expert Syst.* **2021**, *39*, e12781. [CrossRef]

12. Zeinab, K.A.M.; Elmustafa, S.A.A. Internet of things applications, challenges, and related future technologies. *World Sci. News* **2017**, *67*, 126–148.
13. Alqurashi, F.A.; Alsolami, F.; Abdel-Khalek, S.; Saeed, R.A. Machine learning techniques in the internet of UAVs for smart cities applications. *J. Intell. Fuzzy Syst.* **2022**, *42*, 3203–3226. [CrossRef]
14. Tuyishimire, E.; Bagula, A.; Rekhis, S.; Boudriga, N. Trajectory planning for cooperating unmanned aerial vehicles in the IoT. *IoT* **2022**, *3*, 10. [CrossRef]
15. Abdalla, R.S.; Mahbub, S.A.; Mokhtar, R.A.; Ali, E.S.; Saeed, R.A. IoE Design Principles and Architecture. In *Internet of Energy for Smart Cities*; CRC Press: Boca Raton, FL, USA, 2021; pp. 145–170.
16. Ofli, F.; Chaudhry, R.; Kurillo, G.; Vidal, R.; Bajcsy, R. Berkeley MHAD: A comprehensive Multimodal Human Action Database. In Proceedings of the 2013 IEEE Workshop on Applications of Computer Vision (WACV), Clearwater Beach, FL, USA, 15–17 January 2013; pp. 53–60. [CrossRef]
17. Chen, C.; Jafari, R.; Kehtarnavaz, N. UTD-MHAD: A Multimodal Dataset for Human Action Recognition Utilizing A Depth Camera And A Wearable Inertial Sensor. In Proceedings of the 2015 IEEE International Conference on Image Processing, Quebec City, QC, Canada, 27–30 September 2015; pp. 168–172.
18. Hassan, M.B.; Ali, E.S.; Nurelmadina, N.; Saeed, R.A. Artificial intelligence in IoT and its applications. In *Intelligent Wireless Communications*; IET: Stevenage, UK, 2020.
19. Zhang, Y.-D.; Dong, Z.; Gorriz, J.M.; Cattani, C.; Yang, M. Introduction to the Special Issue on Recent Advances on Deep Learning for Medical Signal Analysis. *CMES-Comput. Model. Eng. Sci.* **2021**, *128*, 399–401. [CrossRef]
20. Guan, S.; Zhang, Y.; Tian, Z. Research on Human Behavior Recognition based on Deep Neural Network. *Adv. Comput. Sci. Res.* **2019**, *87*, 777–781.
21. Rashmi, M.; Ram, M.R. Guddeti, Skeleton Based Human Action Recognition for Smart City Application Using Deep Learning. In Proceedings of the 12th International Conference on Communication Systems & Networks (COMSNETS), Bangalore, India, 7–11 January 2020.
22. Thakur, N.; Han, C.Y. An Ambient Intelligence-Based Human Behavior Monitoring Framework for Ubiquitous Environments. *Information* **2021**, *12*, 81. [CrossRef]
23. Liu, X. Sports Deep Learning Method Based on Cognitive Human Behavior Recognition. *Hindawi Comput. Intell. Neurosci.* **2022**, *2022*, 2913507. [CrossRef] [PubMed]
24. Ingle, P.Y.; Kim, Y.-G. Real-Time Abnormal Object Detection for Video Surveillance in Smart Cities. *Sensors* **2022**, *22*, 3862. [CrossRef] [PubMed]
25. Anagnostopoulos, C.-N.E.; Anagnostopoulos, I.E.; Psoroulas, I.D.; Loumos, V.; Kayafas, E. License Plate Recognition From Still Images And Video Sequences: A Survey. *IEEE Trans. Intell. Transp. Syst.* **2008**, *9*, 377–391. [CrossRef]
26. Nurelmadina, N.; Hasan, M.K.; Memon, I.; Saeed, R.A.; Zainol Ariffin, K.A.; Ali, E.S.; Mokhtar, R.A.; Islam, S.; Hossain, E.; Hassan, M.A.; et al. Systematic Review on Cognitive Radio in Low Power Wide Area Network for Industrial IoT Applications. *Sustainability* **2021**, *13*, 338. [CrossRef]
27. Dai, C.; Liu, X.; Lai, J.; Li, P.; Chao, H.-C. Human Behavior Deep Recognition Architecture for Smart City Applications in the 5G Environment. *IEEE Netw.* **2019**, *33*, 206–211. [CrossRef]
28. Hurbean, L.; Danaiata, D.; Militaru, F.; Dodea, A.-M.; Negovan, A.-M. Open Data Based Machine Learning Applications in Smart Cities: A Systematic Literature Review. *Electronics* **2021**, *10*, 2997. [CrossRef]
29. Guerrero-Ibáñez, J.; Zeadally, S.; Contreras-Castillo, J. Sensor technologies for intelligent transportation systems. *Sensors* **2018**, *18*, 1212. [CrossRef] [PubMed]
30. De Las Heras, A.; Luque-Sendra, A.; Zamora-Polo, F. Machine learning technologies for sustainability in smart cities in the post-covid era. *Sustainability* **2020**, *12*, 9320. [CrossRef]
31. Mukhtar, A.M.; Saeed, R.A.; Ali, E.S.; Alhumyani, H. Performance Evaluation of Downlink Coordinated Multipoint Joint Transmission under Heavy IoT Traffic Load. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 6837780. [CrossRef]
32. Alatabani, L.E.; Ali, E.S.; Saeed, R.A. Deep learning approaches for IoV applications and services. In *Intelligent Technologies for Internet of Vehicles*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 253–291.
33. Hassan, M.B.; Ahmed, E.S.A.; Saeed, R.A. Machine Learning for Industrial IoT Systems. In *Handbook of Research on Innovations and Applications of AI, IoT, and Cognitive Technologies*; IGI Global: Hershey, PA, USA, 2021; pp. 336–358.
34. Elfatih, N.M.; Hasan, M.K.; Kamal, Z.; Gupta, D.; Saeed, R.A.; Ali, E.S.; Hosain, M.S. Internet of vehicle's resource management in 5G networks using AI technologies: Current status and trends. *IET Commun.* **2022**, *16*, 400–420. [CrossRef]
35. Ghorpade, S.N.; Zennaro, M.; Chaudhari, B.S.; Saeed, R.A.; Alhumyani, H.; Abdel-Khalek, S. Enhanced differential crossover and quantum particle swarm optimization for IoT applications. *IEEE Access* **2021**, *9*, 93831–93846. [CrossRef]
36. Ahmed, E.S.A.; Mohammed, Z.T.; Hassan, M.B.; Saeed, R.A. Algorithms Optimization for Intelligent IoV Applications. In *Handbook of Research on Innovations and Applications of AI, IoT, and Cognitive Technologies*; IGI Global: Hershey, PA, USA, 2021; pp. 1–25.
37. Wei, H.; Chopada, P.; Kehtarnavaz, N. C-MHAD: Continuous Multimodal Human Action Dataset of Simultaneous Video and Inertial Sensing. *Sensors* **2020**, *20*, 2905. [CrossRef]

38. Alnazir, A.; Mokhtar, R.A.; Alhumyani, H.; Saeed, R.A.; Abdel-khalek, S. Quality of Services Based on Intelligent IoT WLAN MAC Protocol Dynamic Real-Time Applications in Smart Cities. *Comput. Intell. Neurosci.* **2021**, *2021*, 2287531. [CrossRef]
39. Ahmed, Z.E.; Hasan, M.K.; Saeed, R.A.; Hassan, R.; Islam, S.; Mokhtar, R.A.; Khan, S.; Akhtaruzzaman, M. Optimizing Energy Consumption for Cloud Internet of Things. *Front. Phys.* **2020**, *8*, 358. [CrossRef]